

Facile Recovery of Individual High-Molecular-Weight, Low-Copy-Number Natural Plasmids for Genomic Sequencing†

Laura E. Williams,¹ Chris Detter,² Kerrie Barry,² Alla Lapidus,² and Anne O. Summers^{1*}

Department of Microbiology, University of Georgia, Athens, Georgia,¹ and U.S. Department of Energy, Joint Genome Institute, Walnut Creek, California²

Received 13 February 2006/Accepted 3 May 2006

Sequencing of the large (>50 kb), low-copy-number (<5 per cell) plasmids that mediate horizontal gene transfer has been hindered by the difficulty and expense of isolating DNA from individual plasmids of this class. We report here that a kit method previously devised for purification of bacterial artificial chromosomes (BACs) can be adapted for effective preparation of individual plasmids up to 220 kb from wild gram-negative and gram-positive bacteria. Individual plasmid DNA recovered from less than 10 ml of *Escherichia coli*, *Staphylococcus*, and *Corynebacterium* cultures was of sufficient quantity and quality for construction of high-coverage libraries, as shown by sequencing five native plasmids ranging in size from 30 kb to 94 kb. We also report recommendations for vector screening to optimize plasmid sequence assembly, preliminary annotation of novel plasmid genomes, and insights on mobile genetic element biology derived from these sequences. Adaptation of this BAC method for large plasmid isolation removes one major technical hurdle to expanding our knowledge of the natural plasmid gene pool.

Recent genome sequencing and analysis has revealed extensive horizontal gene transfer among bacterial genomes (4, 14). Remnants of mobile genetic elements (MGEs), such as plasmids and bacteriophages (23), are often found adjacent to horizontally transferred chromosomal regions, indicating that these elements are important mediators of gene transfer between bacterial chromosomes. The MGEs themselves also typically carry genes for virulence factors, antibiotic resistances, and novel metabolic processes that enable bacterial hosts to adapt to new environmental conditions (11, 12).

Despite the recognized importance of these elements, genomic analysis of MGEs has been limited. Whereas the total size of sequenced bacterial genomes is 1.3 Gb, only 61 Mb of plasmid genomes and 30 Mb of phage genomes have been sequenced previously (11). Most MGE sequences have been obtained fortuitously during sequencing of their hosts' genomes, resulting in a bias towards MGEs associated with a limited selection of organisms. Large (>50 kb), conjugative plasmids are especially poorly represented in current sequence databases, constituting only 20% of all plasmid sequences in GenBank at present. Most commercial kits for plasmid DNA preparation are designed for small, high-copy-number plasmids. The traditional methods of high-molecular-weight plasmid isolation, such as cesium chloride density gradient centrifugation (26) and pulsed-field gel electrophoresis, require equipment and expertise that are not widely available. Moreover, these and other techniques, such as Eckhardt in-well lysis (6), are time and labor intensive and thus unsuitable for a high-throughput approach.

In the course of large-scale sequencing projects such as the

Human Genome Project, magnetic beads were modified for purification of nucleic acids, including bacterial artificial chromosome (BAC) clones (7, 15, 28). The method used here, termed Solid-Phase Reversible Immobilization (SPRI), employs magnetic beads with carboxylated surfaces to bind plasmid DNA under proprietary buffer conditions. Magnetic immobilization of the beads and bound DNA allows removal of cellular debris and chromosomal DNA. We inferred that SPRI should enable isolation of large, natural plasmids similar in size and copy number to BAC clones. The rapidity, ease, and low cost of SPRI BAC purification suggested that it might provide an advantage over traditional costly and laborious methods of high-molecular-weight plasmid isolation. However, there are some important differences between BACs and natural plasmids.

Whereas BACs are maintained individually in laboratory strains of *Escherichia coli*, wild bacteria typically have several plasmids in a wide range of sizes and copy numbers. Ideally, these should each be recovered separately because the abundance of repetitive elements in plasmids can make computer assembly of libraries constructed from pooled supercoiled DNA, such as obtained from CsCl gradients, difficult or impossible. In addition, it is preferable to recover plasmids from their native hosts (when culturable) rather than having to transfer them to a laboratory strain, which might result in changes (see "Plasmid pLEW517" below). Thus, the ideal plasmid isolation method should be applicable to many types of culturable bacteria and not just *E. coli*.

We describe here a protocol for the use of SPRI for rapid, efficient, and inexpensive isolation of sequencing-quality DNA of individual large, low-copy-number plasmids from gram-negative and gram-positive bacterial strains. As proof of the efficacy of this method, we report on the completion and closure of full-length sequences of five natural bacterial plasmids isolated using this protocol: the previously sequenced 94-kb *Shigella flexneri* plasmid NR1; a novel 65-kb *E. coli* plasmid, pLEW517; a novel 52-kb *Staphylococcus* plasmid, pLEW6932;

* Corresponding author. Mailing address: Department of Microbiology, University of Georgia, 527 Biological Sciences Building, Athens, GA 30602. Phone: (706) 542-2669. Fax: (706) 542-6140. E-mail: summers@uga.edu.

† Supplemental material for this article may be found at <http://aem.asm.org>.

TABLE 1. Bacterial strains and plasmids

Species and host strain (reference) ^a	Plasmid (size in kb)	Plasmid source or reference
<i>Escherichia coli</i>		
DU1040 (10)	NR1 ^c (94)	33
J53 ^b	RP4 (60)	24
CB454	R388 (33)	2
W1485 ^b	F (100)	9
	pBR322 (4)	3
517-2H1 (30)	pLEW517 (65)	32
690FNR (32)	pLEW517 (65)	32
<i>Corynebacterium</i> sp.		
L2-79-05 (22)	pLEW279a (35)	This study
	pLEW279b (30)	This study
<i>Staphylococcus</i> sp.		
693-2 (22)	pLEW6932 (52)	This study

^a The *E. coli* strains are gram negative, the *Corynebacterium* sp. strain is high-G+C gram positive, and the *Staphylococcus* sp. strain is low-G+C gram positive.

^b Strains are deposited in the *E. coli* Genetic Stock Center (<http://cgsc.biology.yale.edu>).

^c Plasmid NR1 is also referred to as R100.

and two novel *Corynebacterium* plasmids, the 35-kb pLEW279a and the 30-kb pLEW279b. We chose these strains and plasmids to answer the following salient questions about plasmid DNA isolation using SPRI technology: (i) does the SPRI isolation method work on gram-negative and gram-positive bacteria? (ii) What is the yield of DNA from a single plasmid band? (iii) What is the best way to remove the plasmid band from the gel? (iv) How much do chromosomal DNA and DNA from other plasmids contaminate the library? We also addressed some of the bioinformatics problems unique to sequencing plasmid DNA: (i) what are the best strategies for vector screening during assembly? (ii) How effective are default BLAST algorithms in identifying DNA and/or protein sequences in plasmids?

MATERIALS AND METHODS

Plasmid DNA isolation. Bacterial strains and plasmids are listed in Table 1. Overnight cultures were inoculated from -70°C stocks and grown without antibiotics. Gram-negative strains were grown overnight in Luria-Bertani (LB) broth at 37°C with shaking at 225 rpm and then diluted 1:50 in LB broth and incubated at 37°C with shaking at 225 rpm for 2.5 to 3 h to obtain cells in logarithmic phase for more effective lysis. Cells were harvested from 1.5 ml culture by centrifugation for 8 min at $11,750 \times g$. Gram-positive strains were grown overnight in brain heart infusion broth at 30°C with shaking at 225 rpm and then diluted in brain heart infusion broth to an optical density at 600 nm of 0.25 to 0.65 to prevent oversaturation of the beads. Cells were harvested from 1.5 ml culture as described above. Gram-positive strains were not subcultured, because the stationary-phase cells lysed as efficiently as cells in logarithmic phase.

Supercoiled DNA was prepared using a CosMCPrep kit for high- and low-copy-number plasmid purification (Agencourt Biosciences Corp., Beverly, MA) and a microcentrifuge tube protocol supplied by the manufacturer. Pelleted cells were suspended in 100 μl CosMCPrep resuspension buffer. Modifications were made to the cell resuspension step for gram-positive organisms to improve lysis. For *Staphylococcus* preparations, resuspension buffer was supplemented with 200 $\mu\text{g/ml}$ lysozyme (Sigma, Inc., St. Louis, MO) and 6% polyethylene glycol (Sigma, Inc.), and cell suspensions were incubated at room temperature for 5 min. For *Corynebacterium* preparations, resuspension buffer was supplemented with 5 mg/ml lysozyme (Sigma, Inc.) and 6% polyethylene glycol, and cell suspensions were incubated at 37°C for 30 min.

Following cell resuspension, 100 μl CosMCPrep lysis buffer was added. Preparations were mixed gently by inverting five times and held at room temperature

for 5 min. All preparations were handled gently during and after lysis to prevent shearing of the supercoiled DNA. CosMCPrep neutralization buffer (100 μl) was added, and preparations were rotated on an orbital shaker for 10 min at 170 rpm to flocculate cellular debris.

Preparations were then centrifuged for 12 min at $16,000 \times g$ at room temperature. The cleared cell lysates (200 μl) were transferred to microcentrifuge tubes and mixed with 139 μl isopropanol and 10 μl CosMCPrep magnetic bead suspension by mixing gently four to six times with a pipette tip. Tubes were placed in a magnetic stand at room temperature for 10 min to trap the beads with their bound DNA against the sides of the tubes. Unless otherwise noted, all subsequent steps were conducted at room temperature with the tubes in the magnetic stand. The lysate was aspirated without disturbing the beads, and the beads were washed three times with 200 μl 70% ethanol–30% autoclaved MilliQ water. Beads were dried at 37°C in a forced air incubator for 2 min. DNA was then eluted by pipetting 40 μl CosMCPrep resuspension buffer over the beads. Preparations were returned to 37°C in closed tubes for 5 min to separate any beads from the eluate and allow complete elution of bound supercoiled DNA. The bead-free supercoiled DNA eluates (40 μl) were transferred to microcentrifuge tubes and stored at -20°C . Electrophoresis and gel extraction of plasmids were usually done within 24 h, but eluates from gram-negative strains were stored for up to 20 days without appreciable loss of supercoiled DNA.

Electrophoresis and gel extraction of plasmids. All 40 μl of eluate was loaded onto a 0.3% or 0.5% SeaKem Gold agarose (Cambrex BioScience, Walkersville, MD) gel, with 10 μl of High Range MassRuler DNA ladder (Fermentas, Inc., Hanover, MD) as a size and mass standard. The gel was electrophoresed in $1 \times$ TAE (40 mM Tris-acetate, 2 mM $\text{Na}_2\text{EDTA}-2\text{H}_2\text{O}$) at 80 V for 3 to 4 h. Gels were stained with SYBR green I nucleic acid gel stain (Molecular Probes, Inc., Eugene, OR) to increase sensitivity and reduce the DNA damage typically observed with ethidium bromide. Gels were visualized on a Molecular Dynamics FluorImager, and the mass of DNA in each plasmid band was determined by densitometry. To reduce exposure to damaging UV light, gels were placed on a DarkReader Transilluminator (Clare Chemical Research, Denver, CO), and gel slices containing individual plasmid bands were excised with a razor blade. Supercoiled plasmid DNA was extracted from gel slices using either a GeneClean Turbo glass milk spin kit (Qbiogene, Inc., Carlsbad, CA) or dialysis tubing electroelution.

A GeneClean Turbo kit was used following the manufacturer's instructions, except that 30 μl TE buffer (10 mM Tris, 1 mM EDTA, pH 8) was used for elution. Dialysis tubing electroelution was done as previously described (27, 29), using 1-inch-diameter SpectraPor dialysis tubing that had been boiled in 25 mM EDTA for 10 min, rinsed once with water, and stored at 4°C in 30% ethanol. Gel slices with plasmid DNA were placed in dialysis tubing bags with 250 to 500 μl of $0.1 \times$ TAE and electrophoresed at 100 V for 2 h in $0.1 \times$ TAE. Polarity was reversed for 2 min, and then the electroeluted DNA was transferred by pipette to a clean 1.5 ml microcentrifuge tube.

DNA yield was quantified by $A_{260/280}$ of a 1:10 dilution in either TE (for GeneClean eluates) or $0.1 \times$ TAE (for dialysis tubing electroeluates).

Whole-genome shotgun sequencing and assembly. For each plasmid, four preparations of DNA, each isolated from 1.5 ml culture using the appropriate CosMCPrep kit method, excised from a gel as a single plasmid band, and eluted using either the GeneClean kit or dialysis tubing electroelution, were pooled and then sheared on a HydroShear (Genomic Solutions, Ann Arbor, MI) at setting 9 into 2- to 3-kb fragments, which were blunt end ligated into pMCL200, a pUC18-based cloning vector (21). Library construction and template preparation were conducted following standard Joint Genome Institute protocols (http://www.jgi.doe.gov/sequencing/protocols/protos_production.html). End sequencing reactions were carried out using a 1/16 dilution of BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA) and resolved on ABI PRISM 3730 sequencers. Electropherograms were analyzed with PHRED basecalling software (8). The average sequencing read length was 689 ± 30 bp. Sequencing reads were screened using Cross-Match SPS-3.57 (Southwest Parallel Software) to identify and remove vector sequence. For the *Corynebacterium* plasmids pLEW279a and pLEW279b, the entire vector sequence was used for screening. For the *E. coli* plasmids NR1 and pLEW517 and the *Staphylococcus* plasmid pLEW6932, screening with the complete cloning vector sequence introduced artificial gaps into the assemblies, possibly due to similarity between the origins of replication and antibiotic resistance genes on the natural plasmids and those on the cloning vector. Consequently, only sequences identical to the cloning vector spanning the insert site to the sequencing primer annealing site were removed from sequencing reads for these plasmids. After the vector sequences were removed, reads were assembled by PHRAP (www.phrap.org), and gap closure was accomplished by directed PCR of library clones or purified plasmid DNA, resulting in a single, circularized contiguous sequence (contig) for each plasmid.

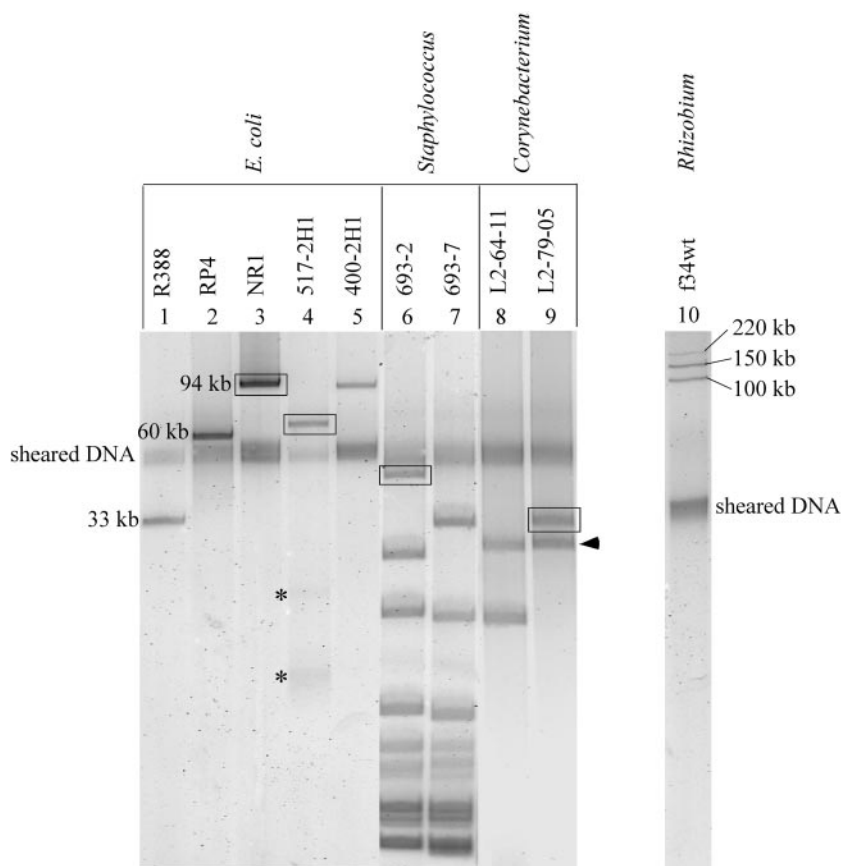


FIG. 1. Plasmid DNA prepared using magnetic bead-based SPRI. Lanes 1 to 9, 0.5% SeaKem Gold agarose-1× TAE. Lane 1 = CB454 (R388). Lane 2 = J53 (RP4). Lane 3 = DU1040 (NR1). Lanes 4 and 5 = wild *E. coli* strains (30). Lanes 6 and 7 = wild *Staphylococcus* strains (22). Lanes 8 and 9 = wild *Corynebacterium* strains (22). Lane 10 = *Rhizobium meliloti*, 0.75% Sigma agarose-1× TBE, run at 80 V for 8 h. The sheared DNA band contains any linearized DNA larger than 20 kb. Boxes indicate plasmids whose sequencing is reported here. Asterisks indicate additional faint plasmid bands in *E. coli* strain 517-2H1, and an arrowhead indicates the plasmid whose sequence was obtained adventitiously along with that of the boxed plasmid of *Corynebacterium* strain L2-79-05.

Analysis of plasmid sequences. BLAST (1) analysis used the default parameters for the NCBI BLASTN program (+1/−3 match/mismatch penalty; word size 11) and the BLASTX program (word size 3), with the exception that the bacterial translation table was used for BLASTX (www.ncbi.nlm.nih.gov/BLAST/). Pairwise, whole genome alignments of plasmid sequences were generated using the MUMmer software package (19). The “nucmer” script was used to obtain percent identity between the two sequences and to identify the exact locations of disagreements.

GenBank accession numbers. The plasmid sequences determined in this study have been deposited in the GenBank database under the following accession numbers: for NR1, DQ364638; for wild pLEW517, DQ390454; for transconjugant pLEW517, DQ390455; for pLEW6932, DQ390456; for pLEW279a, DQ390458; and for pLEW279b, DQ390457.

RESULTS

Recovery of supercoiled plasmid DNA. The standard CosMCPrep SPRI protocol was effective for plasmid isolation from gram-negative proteobacteria *E. coli*, *Rhizobium* (Fig. 1), *Salmonella*, and *Pseudomonas* (data not shown), and modifications (see Materials and Methods) enabled plasmid isolation from low-G+C (*Staphylococcus*) and high-G+C (*Corynebacterium*) gram-positive bacteria (Fig. 1). Plasmids from 4 to 220 kb in size were readily visible on agarose gels, and supercoiled DNA was recovered equally well from small, high-copy-num-

ber plasmids such as pBR322 (3) and from large, low-copy-number plasmids such as NR1 (33) (Fig. 1). For the well-characterized, laboratory standard plasmids pBR322, R388, RP4, and NR1, the average yield of DNA of each plasmid recovered from 1.5 ml *E. coli* cultures and detected as a single band on a gel was 0.865 µg (range, 0.5 to 1.3 µg). Host-plasmid combinations were chosen to address specific critical questions concerning the ability of this method to optimize the sequencing of large, low-copy-number natural plasmids directly from their original host bacteria.

Plasmid NR1. To determine whether the recovered plasmid DNA was suitable in amount and quality for library construction, we sequenced plasmid NR1, a 94-kb Inc FII plasmid presumably identical to the previously sequenced plasmid R100 (NC_002134). We found that using a GeneClean kit for gel extraction of NR1 DNA resulted in extensive random shearing (Fig. 2). However, pooling four preparations yielded 1.1 µg DNA, which was sufficient for construction of a library of 768 clones without amplification. Sequencing and assembly of NR1 yielded one contig of 94,289 bp and another of 2,198 bp (Table 2). MUMmer (19) analysis showed that the larger contig had only 15 single-nucleotide disagreements with the

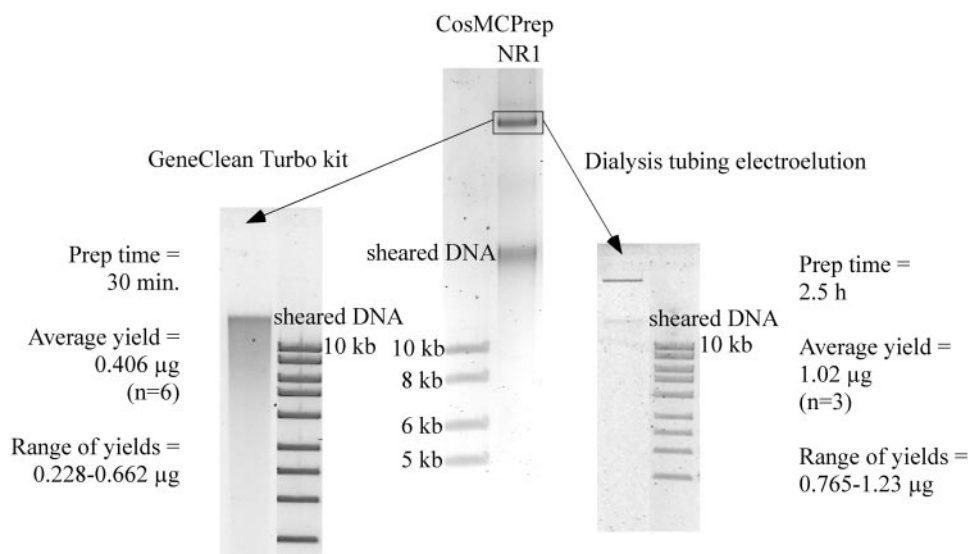


FIG. 2. Extraction of NR1 plasmid DNA from agarose gels. A 10 μ l aliquot of the extracted DNA was run on a 0.75% Sigma agarose-1 \times TBE gel.

94,281-bp R100 reference sequence (see supplemental material published online). The sequence quality for all reads was sufficiently high to confirm that these were real polymorphisms in the plasmid sequences.

The smaller contig from the NR1 project (Table 2) was 100% identical to the pMCL200 cloning vector and was likely a result of the modified vector screening used during assembly. The narrowed vector screening was required to prevent introduction of artificial gaps in the NR1 assembly in regions of identity between the natural plasmid and the cloning vector (e.g., *cat*, the chloramphenicol acetyltransferase gene). However, allowing some vector sequences to remain resulted in the assembly of the small but readily identifiable vector contig.

Plasmid pLEW517. To determine whether a single plasmid in a multiplasmid host could be reliably sequenced from supercoiled DNA recovered from a gel, we sequenced plasmid pLEW517, previously shown to confer resistance to ampicillin, streptomycin, sulfonamides, and mercury (32), as derived both from its native host, the multiplasmid primate intestinal *E. coli* strain 517-2H1 (Fig. 1) (30), and from an otherwise plasmid-free laboratory *E. coli* strain 690FNR into which it had been transferred by conjugation. Dialysis tubing electroelution,

which preserved the supercoiled conformation of extracted DNA better than glass milk extraction (Fig. 2), was used to recover pLEW517 DNA and in all subsequent work. Pooling four preparations yielded 3.5 μ g for wild pLEW517 and 4.0 μ g for transconjugant pLEW517, amounts which were sufficient for library construction.

Sequencing and assembly of wild pLEW517 yielded a single major contig of 63,946 bp (Table 2), and transconjugant pLEW517 yielded a single major contig of 65,288 bp. These contigs were 100% identical except for a 1,342-bp segment found on transconjugant pLEW517 but not wild pLEW517 (discussed below). Overall, 98% of the sequence of pLEW517 returned significant BLASTN hits (Table 3). BLASTN analysis indicated that pLEW517 is a variant of plasmid R46 (NC_003292), a 50-kb IncN plasmid previously observed in *Salmonella enterica* serovar Typhimurium. Significant similarity (e-value, 0.0) was detected to regions encoding replication, maintenance, and transfer functions on plasmid R46. However, an 18-kb segment of R46 that included the *InI* integron was absent from pLEW517. pLEW517 also had sequences with very significant similarity to those of transposon Tn21, which contains a class 1 integron (*In2*) and a mercury resistance (*mer*) operon (20), and of transposon Tn3, which encodes a β -lactamase (18).

Alignments of the wild and transconjugant pLEW517 sequences to the R46 sequence indicated that the 1,342-bp difference between the two sequences lies in a repeat region of conserved upstream (CUP) elements, which may regulate expression of adjacent genes during conjugative transfer (5). Repeat regions frequently present difficulties during assembly, and this repeat region may have been overcollapsed during assembly of the wild pLEW517 sequence, thus causing the observed difference in length. However, the high sequencing quality and coverage suggests that this may be a real difference between the wild and transconjugant plasmids arising during conjugation to the new host. In either case, neither chromosomal DNA nor the two other plasmids of very different molecular weights and copy numbers in the same strain (Fig. 1)

TABLE 2. Sequencing coverage resulting in single, circularized contigs

Plasmid and type	Contig	Length (bp)	No. of sequencing reads	Depth (\times)
De novo NR1	1	94,289	1,141	11
	2	2,198	306	108
pLEW517				
Wild	1	63,946	1,243	13
Transconjugant	1	65,288	1,271	15
pLEW6932	1	51,514	2,648	44
pLEW279a	1	34,606	3,759	90
pLEW279b	1	29,854	303	9

TABLE 3. Highlights of the chimeric character of novel plasmid genomes^a

Plasmid and BLAST program	Selected BLAST hit ^b (% identity)	% of plasmid genome similar to hit	Gene(s) and/or region(s) of interest within hit
pLEW517			
BLASTN	<i>S. enterica</i> serovar Typhimurium plasmid R46 (99% nt)	49.1	<i>oriT</i> , Tra proteins, <i>repA</i> , iterons, <i>korAB</i>
	Tn21 (99% nt)	18.3	<i>mer</i> operon, class 1 integron, <i>tnpA</i>
	Tn3 (97% nt)	7.7	β -Lactamase
pLEW6932			
BLASTN	<i>Staphylococcus saprophyticus</i> pSSP1 (97% nt)	19.6	Arsenic resistance operon
	<i>S. aureus</i> MSSA476 (84% nt)	6.7	Metallo- β -lactamase superfamily protein
BLASTX	<i>Staphylococcus</i> plasmids (50–65% aa)	1.8	RepA
	Multiple species (30–40% aa)	1.2	Cation transport ATPase
	Multiple species (50–70% aa)	1.0	Glycine betaine transporter
pLEW279a			
BLASTN	<i>C. glutamicum</i> pTET3 (99% nt)	30.6	IS6100, <i>tetA</i> , <i>tetR</i> , class 1 integron w/ truncated <i>intI1</i>
	<i>Arcanobacterium pyogenes</i> pAP2 (99% nt)	8.1	<i>ermX</i>
BLASTX	<i>C. glutamicum</i> pTET3 and pCG4 (60% aa)	4.1	RepA
	<i>C. diphtheriae</i> pNG2, <i>C. glutamicum</i> pGA2 (50% aa)	9.6	TraA
pLEW279b			
BLASTN	<i>C. glutamicum</i> ATCC 13032 chromosomal DNA (97% nt)	22	Two-component system response regulator and kinase, Cu ²⁺ transport ATPase, Cu ²⁺ oxidase
	<i>C. jeikeium</i> pA505 (99% nt)	6.6	Putative ABC transporter
BLASTX	<i>C. striatum</i> pTP10 (69% aa)	3.1	RepA
	<i>Rhodococcus equi</i> p103 (31% aa)	9.9	TraA-like transfer protein

^a Data do not represent an exhaustive list of BLAST hits for these plasmids.^b Where appropriate, multiple hits with similar BLAST scores are listed.

interfered with the assembly of the target plasmid extracted from the gel.

Further examination of the pLEW517 region showing similarity to Tn21 revealed that although the hallmarks of Tn21, which include the transposition genes (*tnpAR*) and the *mer* operon, were present on the pLEW517 transposon, there were dramatic differences in the content of its version of the integron (Fig. 3). Whereas the class 1 integron In2 of Tn21 has a single cassette encoding an aminoglycoside adenylyltransferase (*aadA1*), the class 1 integron of pLEW517 has three cassettes carrying *dfrA12* (previously called *dhfrXII*), a dihydrofolate reductase; an open reading frame (ORF) of unknown function; and *aadA2*, an aminoglycoside adenylyltransferase. These cassettes were identified by BLASTN analysis based on similarity to an integron found on the 89.5-kb *Citrobacter freundii* plasmid pCTX-M3 (NC_004464), which contains the three cassettes in the same order. This cassette arrangement was previously reported on a Tn21-like element carried on a 70-kb plasmid from a pathogenic *E. coli* strain (16). Note that the insertion of the pLEW517 integron into the ancestral *mer* transposon “backbone” is in exactly the same position as in the prototypical Tn21 of NR1 and R100.

In addition to these differences at the *attI* insertion point, there was an insertion into the pLEW517 integron that truncated both *orf5* in the 3' conserved segment and the transposition (*tni*) module compared to the corresponding sequences in the prototypical Tn21. BLAST identified similarity to a region on the 48-kb plasmid pRSB101 (NC_006385) from an uncultured host, including a homolog of *chrA* resembling a chromate ion transporter, an ORF of unknown function, an

operon carrying a macrolide phosphotransferase (*mphA*) and its repressor (*mphR*), and two insertion sequence (IS) elements. The macrolide resistance operon was previously observed inserted into the integron of a Tn21-like transposon in a strain of *Aeromonas hydrophila* isolated from swine (25). This integron also contained the *dfrA12*, *orfF*, and *aadA2* cassettes identified on the pLEW517 transposon; however, it lacked the *chrA* homolog, the IS elements, and the ORF of unknown function. To our knowledge, this is the first observation of a Tn21-like transposon in an IncN plasmid, demonstrating that Tn21 is not limited to the R100-like IncFII plasmids with which it is commonly associated. The differences in accessory element content between the prototypical Tn21, the other Tn21-like elements as described above, and the pLEW517 transposon support the idea that Tn21-like elements and their host plasmids serve as hotspots for recombination involving integron cassettes, insertion sequences, and other transposons.

Plasmid pLEW6932. We then assessed the effectiveness of the SPRI method on low-G+C gram-positive bacteria using the multiplasmid *Staphylococcus* strain 693-2 obtained from poultry litter (Fig. 1) (22). This strain had nine visible plasmid bands, the largest of which was chosen for sequencing. Four pooled preparations yielded 2.9 μ g DNA for library construction. Sequencing and assembly produced a major contig of 51,514 bp, similar to the 51 kb estimated for pLEW6932 from agarose gels. BLASTN identified two small regions of significant similarity (e-value, 0.0): one similar to the arsenic resistance operon of *Staphylococcus saprophyticus* plasmid pSSP1 (NC_007351) and another to β -lactamase genes from other *Staphylococcus* chromosomes and plasmids (e.g., BX571857)

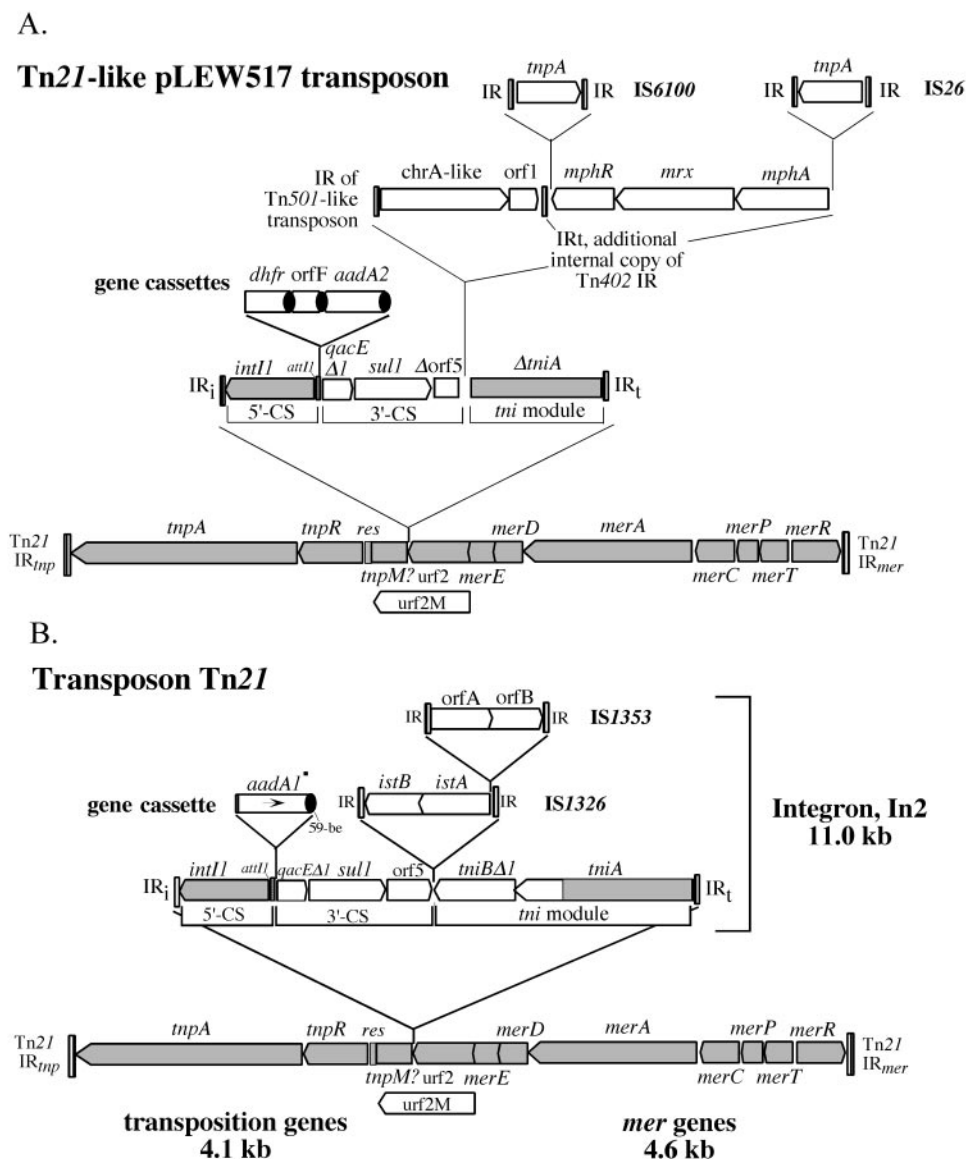


FIG. 3. (A) The Tn21-like transposon found on *E. coli* plasmid pLEW517. Shading indicates regions where the pLEW517 transposon shares highest similarity with Tn21. Inverted repeats (IR) of transposons and insertion sequences are indicated by vertical bars. The transposition genes include the transposase (*tnpA*), the resolvase (*tnpR*), and the putative transposition regulator (*tnpM*). The transposition resolution site is encoded by *res*. The 5' conserved segment (CS) of the integrin consists of the integrase (*intI1*) and the *aatI1* recombination site. Three gene cassettes carry dihydrofolate reductase (*dhfr*), an ORF of unknown function (*orfF*), and aminoglycoside adenylyltransferase (*aadA2*). The 3' conserved segment (CS) consists of genes encoding resistance to quaternary ammonium compounds (*qacEΔI*), sulfonamide resistance (*sulI*), and an ORF of unknown function (*orf5*), which is truncated compared to *orf5* in Tn21. An additional element has inserted into the integrin. This element contains a *chrA*-like gene resembling a chromate ion transporter, an ORF of unknown function (*orf1*), genes encoding a macrolide phosphotransferase (*mphA*) and its repressor (*mphR*), a gene encoding a hydrophobic protein (*mrX*), and two insertion sequence elements (*IS6100* and *IS26*). The integrin transposition (*tni*) module of this transposon has also been truncated compared to Tn21, with only a truncated *tniA* remaining. The mercury resistance operon includes regulatory genes (*merR* and *merD*) and structural genes (*merT*, *merP*, *merC*, and *merA*). (B) Tn21 as depicted in reference 20. Figure 3A was adapted from reference 20.

(Table 3). The arsenic resistance (*ars*) operon on pLEW6932 may be selected due to the common use of organic arsenic coccidiostats such as roxarsone for growth promotion; roxarsone naturally degrades to inorganic arsenate and arsenite (13), to which this locus confers resistance.

To look for genes conserved only at the amino acid sequence level, we used BLASTX analysis, in which the nucleotide sequence is translated into all six possible reading frames and com-

pared to the protein sequence database. BLASTX identified a region of significant similarity ($e\text{-value} \leq e^{-90}$) to replication initiation proteins of various *Staphylococcus* plasmids (e.g., NP_932180, CAA63141). Other regions of similarity included hits to glycine betaine transporters (e.g., ZP_00233406) and cation transport ATPases (e.g., ZP_00063375) found on the chromosomes of a variety of bacterial species. Of the 125 hits identified using default parameters of BLASTX, most were repetitive hits

on these and a few other loci, leaving approximately 60% of the pLEW6932 sequence with no known protein or nucleic acid similarities as reported with the default parameters of these BLAST programs.

Plasmids pLEW279a and pLEW279b. Last, *Corynebacterium* strain L2-79-05 provided the opportunity to assess the effectiveness of SPRI on high-G+C gram-positive bacteria and also to investigate the occurrence of plasmid band cross-contamination in multiplasmid strains. In contrast to *E. coli* 517-2H1, the plasmid profile of this poultry litter (22) strain shows two plasmids of very similar sizes and copy numbers (Fig. 1). Isolation of the larger plasmid pLEW279a by electroelution from the gel yielded 3.1 µg DNA for library construction. Sequencing and assembly produced two major closed circular contigs (Table 2). The 34,606-bp contig corresponded in size to the 35-kb band extracted from agarose gels, and the second contig of 29,854 bp was similar in size to the 30-kb plasmid (named pLEW279b) also observed in the L2-79-05 plasmid profile. This suggests that for plasmids of very similar sizes and copy numbers, gel slices of apparently single-plasmid bands may contain some of the other plasmid's DNA; however, current methods of sequence assembly are capable of resolving the two plasmids into separate contigs. Although the fortuitously sequenced pLEW279b has fewer sequencing reads than pLEW279a (Table 2), it still has very good depth.

For pLEW279a, BLASTN identified two regions with significant similarity (e-value, 0.0) to other plasmids. One large region resembles part of the 28-kb *Corynebacterium glutamicum* plasmid pTET3 (NC_003227) (31), including a tetracycline resistance determinant and a class 1 integron with a truncated integrase and an aminoglycoside adenylyltransferase cassette, *aadA9* (Table 3). An adjacent, smaller region resembles the 9-kb *Arcanobacterium pyogenes* plasmid pAP2 (NC_005206) (17), including a macrolide resistance determinant. In addition, BLASTX identified a region of pLEW279a with approximately 60% amino acid identity to RepA replication initiation proteins of *C. glutamicum* plasmids pTET3 and pCG4 (29 kb; NC_004945) and a region with approximately 50% amino acid identity to TraA transfer proteins of *Corynebacterium* plasmids pGA2 (19 kb; NC_004535) and pNG2 (15 kb; NC_005001). Of the 148 BLASTX hits, all are repeated hits on these and a few other loci, leaving 40% of the plasmid sequence with no known protein or nucleic acid similarities as reported by the default parameters of the BLAST programs.

For pLEW279b, BLASTN identified a large region of significant similarity (e-value, 0.0) to *C. glutamicum* ATCC 13032 chromosomal DNA (BA000036) containing genes involved in copper metabolism and genes encoding a two-component-type response regulator and kinase. BLASTN also identified two small regions of similarity to *Corynebacterium* plasmids: one region of approximately 2 kb similar to a putative ABC transporter found on the 12-kb *C. jeikeium* plasmid pA505 (NC_004773) and another region of approximately 0.5 kb similar to part of a transposase found on *C. glutamicum* plasmids pTET3, pAG1 (20 kb; NC_001415), and pGA2. Last, BLASTX identified a region with approximately 69% amino acid identity to the RepA replication protein of *C. striatum* plasmid pTP10 (51.5 kb; NC_004939) and a region with approximately 31% amino acid identity to the TraA-like transfer protein of *Rhodococcus equi* virulence plasmid p103 (80.5 kb; NC_002576). The

103 BLASTX hits were repeated instances of these and a few other sequences. Approximately 40% of this plasmid sequence had no known protein or nucleic acid similarities detectable by the default parameters of these two BLAST programs.

DISCUSSION

The SPRI method described here can be combined with lysis methods for other bacteria that do not perturb the attachment of plasmid DNA to the carboxylated microspheres. This approach, in addition to being facile and inexpensive (ca. \$0.50 per preparation), enables direct recovery of plasmid DNA from wild bacterial strains without the need to transfer them to a laboratory strain of *E. coli*. The yield of plasmid DNA after isolation, electrophoresis, and elution is sufficient in that pooling a small number of preparations provides enough sequencing-quality DNA to construct a library and obtain sequence at excellent coverage ($>9\times$). We did not detect any chromosomal contamination of the plasmid sequences, and contamination of plasmid DNA preparations with other plasmid DNA was only detected when the two plasmids had similar sizes and copy numbers. Even in this instance, current methods of sequence assembly were able to differentiate the two sequences.

We found that vector screening during assembly of plasmid sequences is complicated by the presence of similar genes on both the natural plasmid and the cloning vector, resulting in the introduction of artificial gaps in the assembly. Rather than using the entire cloning vector sequence for screening, we screened using only the cloning vector sequence from the insert site to the primer annealing site. This resolved the gaps in the assemblies; however, in one case it resulted in the generation of a small contig identical to the cloning vector's origin of replication and antibiotic resistance marker. Since this contig was easily identified as cloning vector sequence, the strategy of narrowed vector screening was considered suitable for plasmid sequence assembly. As a general rule in plasmid sequencing work, any antibiotic resistance phenotype information available for the plasmid of interest could guide selection of a cloning vector that encodes a different antibiotic resistance than those found on the natural plasmid.

Our preliminary analysis of the four novel plasmid sequences illustrates the variety of genes that can be carried by plasmids as well as the limitations in plasmid genome information currently available. BLAST analysis successfully identified plasmid replication, transfer, and maintenance genes as well as mobile and chromosomal genes with diverse functions in all novel plasmid sequences. The gram-negative plasmid pLEW517, which had high-scoring BLAST hits for all but a small fraction of the genome, was characterized as a variant of plasmid R46. pLEW517 encoded the same replication and transfer genes as R46 but lacked a characteristic R46 integron and possessed two transposons not found on R46. This illustrates how the insertion and removal of accessory elements such as transposons, integrons, and insertion sequences on a plasmid backbone leads to extensive variation in plasmid genomes, similar to the chromosomal variation observed among strains of a bacterial species. Further sequencing of natural plasmids will contribute to our understanding of the degree to which plasmid backbones vary in accessory elements. Sequencing more plasmids will also uncover variation in the accessory

elements themselves, such as the novel Tn21-like transposon found on pLEW517.

In contrast to pLEW517, approximately 60% of the pLEW6932 sequence and 40% of the pLEW279a and pLEW279b sequences had no apparent similarity to known sequences in the nucleotide and protein sequence databases. BLAST hits obtained for these genomes were from a variety of gram-positive bacterial chromosomes and plasmids, suggesting that these plasmid genomes are a mosaic of genes and accessory elements acquired from related sources. The underrepresentation of environmental gram-positive plasmids in current sequence databases may account for the observed lack of similarity. Thorough manual annotation of these plasmids is beyond the expertise of a single laboratory; however, the initial annotation accompanying the sequences as deposited in GenBank can provide a starting point for deeper analysis by those with relevant expertise. Much more extensive sequencing of plasmid genomes as well as inclusion of plasmid and other mobile element terms in the genome and sequence ontologies is essential to addressing fundamental questions of prokaryotic evolution and to dealing with critical problems such as the spread of antibiotic resistance. The SPRI plasmid isolation method coupled with simple electrophoretic elution of individual supercoiled plasmids provides a facile and inexpensive approach to obtain DNA of sufficient amount and quality to generate libraries with excellent coverage that result in readily finishable genomic sequences.

ACKNOWLEDGMENTS

We thank NSF-REU (grant DBI0453353) summer student David Miller for assistance in applying the SPRI method to gram-positive bacteria, colleagues Tamar Barkay, Margie Lee, John Maurer, Mark Schell, and Patricia Sobecky for advice on techniques and critical reading of the manuscript, and three anonymous reviewers for their help in improving the manuscript. We also thank Agencourt, Inc. for providing us with a small CosMCPrep starter kit.

This work was supported by the Office of Science (BER), U.S. Department of Energy (grant DE-FG02-04ER63770 to A.O.S.).

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Avila, P., and F. de la Cruz. 1988. Physical and genetic map of the IncW plasmid R388. *Plasmid* **20**:155–157.
- Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**:95–113.
- Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case, and W. F. Doolittle. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* **37**:283–328.
- Delver, E. P., and A. A. Belogurov. 1997. Organization of the leading region of IncN plasmid pKM101 (R46): a regulation controlled by CUP sequence elements. *J. Mol. Biol.* **271**:13–30.
- Eckhardt, T. 1978. A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. *Plasmid* **1**:584–588.
- Elkin, C. J., P. M. Richardson, H. M. Fourcade, N. M. Hammon, M. J. Pollard, P. F. Predki, T. Glavina, and T. L. Hawkins. 2001. High-throughput plasmid purification for capillary sequencing. *Genome Res.* **11**:1269–1274.
- Ewing, B., L. Hillier, M. C. Wendt, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
- Firth, N., K. Ippen-Ihler, and R. A. Skurray. 1996. Structure and function of the F factor and mechanism of conjugation, p. 2377–2401. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 2. ASM Press, Washington, D.C.
- Foster, T. J., H. Nakahara, A. A. Weiss, and S. Silver. 1979. Transposon A-generated mutations in the mercuric resistance genes of plasmid R100-1. *J. Bacteriol.* **140**:167–181.
- Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint. 2005. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**:722–732.
- Funnell, B. E., and G. J. Phillips (ed.). 2004. *Plasmid biology*. ASM Press, Washington, D.C.
- Garbarino, J. R., A. J. Bednar, D. W. Rutherford, R. S. Beyer, and R. L. Wershaw. 2003. Environmental fate of roxarsone in poultry litter. I. Degradation of roxarsone during composting. *Environ. Sci. Technol.* **37**:1509–1514.
- Gogarten, J. P., and J. P. Townsend. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**:679–687.
- Hawkins, T. L., T. O'Connor-Morin, A. Roy, and C. Santillan. 1994. DNA purification and isolation using a solid-phase. *Nucleic Acids Res.* **22**:4543–4544.
- Heikkila, E., M. Skurnik, L. Sundstrom, and P. Huovinen. 1993. A novel dihydrofolate reductase cassette inserted in an integron borne on a Tn21-like element. *Antimicrob. Agents Chemother.* **37**:1297–1304.
- Jost, B. H., A. C. Field, H. T. Trinh, J. G. Songer, and S. J. Billington. 2003. Tylosin resistance in *Arcanobacterium pyogenes* is encoded by an *ermX* determinant. *Antimicrob. Agents Chemother.* **47**:3519–3524.
- Kostryken, R., C. Morita, and F. Heffron. 1981. Transposon Tn3 encodes a site-specific recombination system: identification of essential sequences, genes, and actual site of recombination. *Proc. Natl. Acad. Sci. USA* **78**:4041–4045.
- Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* **5**:R12.
- Liebert, C. A., R. M. Hall, and A. O. Summers. 1999. Transposon Tn21, flagship of the floating genome. *Microbiol. Mol. Biol. Rev.* **63**:507–522.
- Nakano, Y., Y. Yoshida, Y. Yamashita, and T. Koga. 1995. Construction of a series of pACYC-derived plasmid vectors. *Gene* **162**:157–158.
- Nandi, S., J. J. Maurer, C. Hofacre, and A. O. Summers. 2004. Gram-positive bacteria are a major reservoir of class 1 antibiotic resistance integrons in poultry litter. *Proc. Natl. Acad. Sci. USA* **101**:7118–7122.
- Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299–304.
- Pansegau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich, and C. M. Thomas. 1994. Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J. Mol. Biol.* **239**:623–663.
- Poole, T. L., T. R. Callaway, K. M. Bischoff, C. E. Warnes, and D. J. Nisbet. 2006. Macrolide inactivation gene cluster *mphA-mrx-mphR* adjacent to a class 1 integron in *Aeromonas hydrophila* isolated from a diarrhoeic pig in Oklahoma. *J. Antimicrob. Chemother.* **57**:31–38.
- Radloff, R., W. Bauer, and J. Vinograd. 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc. Natl. Acad. Sci. USA* **57**:1514–1521.
- Sambrook, J., and D. W. Russell. 2001. *Molecular cloning: a laboratory manual*, 3rd ed., vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Skowronski, E. W., N. Armstrong, G. Andersen, M. Macht, and P. M. McCready. 2000. Magnetic, microplate-format plasmid isolation protocol for high-yield, sequencing-grade DNA. *BioTechniques* **29**:786–792.
- Strong, S. J., Y. Ohta, G. W. Litman, and C. T. Amemiya. 1997. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. *Nucleic Acids Res.* **25**:3959–3961.
- Summers, A. O., J. Wireman, M. J. Vimy, F. L. Lorscheider, B. Marshall, S. B. Levy, S. Bennett, and L. Billard. 1993. Mercury released from dental “silver” fillings provokes an increase in mercury- and antibiotic-resistant bacteria in oral and intestinal floras of primates. *Antimicrob. Agents Chemother.* **37**:825–834.
- Tauch, A., S. Gotker, A. Puhler, J. Kalinowski, and G. Thierbach. 2002. The 2.78-kb R-plasmid pTET3 from *Corynebacterium glutamicum* encodes the aminoglycoside adenylyltransferase gene cassette *aadA9* and the regulated tetracycline efflux system Tet 33 flanked by active copies of the widespread insertion sequence IS6100. *Plasmid* **48**:117–129.
- Wireman, J., C. A. Liebert, T. Smith, and A. O. Summers. 1997. Association of mercury resistance with antibiotic resistance in the gram-negative fecal bacteria of primates. *Appl. Environ. Microbiol.* **63**:4494–4503.
- Womble, D. D., and R. H. Rownd. 1988. Genetic and physical map of plasmid NR1: comparison with other IncFII antibiotic resistance plasmids. *Microbiol. Rev.* **52**:433–451.